H. Leppäkoski, S. Tikkinen and J. Takala, "Optimizing radio map for WLAN fingerprinting," 2010 Ubiquitous Positioning Indoor Navigation and Location Based Service, Kirkkonummi, 2010, pp. 1-8.
doi: 10.1109/UPINLBS.2010.5654332
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5654332&isnumber=5648783

Optimizing Radio Map for WLAN Fingerprinting

Helena Leppäkoski, Saija Tikkinen, and JarmoTakala
Department of Computer Systems
Tampere University of Technology
Tampere, Finland
helena.leppakoski@tut.fi

Abstract— **In this paper, questions related to design of WLAN radio map for fingerprinting based positioning were investigated. The experiment results show that with histogram based algorithms, the positioning accuracy improves as the number of histogram bins increases until the number of bins reaches eight. With the number of bins lower than this, the uneven bin distribution with separate bin for missing samples gives better accuracy than even bin widths. If the calibration data contains samples from several measurement directions, it is beneficial to combine them into one fingerprint, as this decreases the size of the radio map and gives at least the same accuracy as having separate fingerprints for different measurement directions. In the experiments, the combination of the signals from correlating sources before the computation of the radio map and position estimate decreases the positioning accuracy only by 1-2 m, but decreases significantly the size of the radio map. The best method for signal combinations depends on the bin configuration and position estimation algorithm.**

Keywords-Position estimation; Received signal strength; Pattern recognition; Bayesian methods; Histograms

I. INTRODUCTION

For positioning methods based on WLAN signals, several possible observables have been proposed: Received Signal Strength (RSS), Time of Arrival (TOA), Time Difference of Arrival (TDOA), or Angle of Arrival (AOA). From these, the RSS observables are the easiest to be applied in consumer market positioning applications since mobile terminals (MT) can obtain them by passive scanning of WLAN beacon frames, which the infrastructure WLAN Access Points (AP) emit periodically according to IEEE 802.11 standard [1]. In addition, in many mobile devices, such as mobile phones, Personal Digital Assistants (PDAs) and laptop computers, RSS measurements are easily available through Application Programming Interfaces (APIs) of their standard WLAN services.

In RSS based positioning, the MT location is estimated by using models that relate the strength of the received radio signal either to the distance between the MT and the signal emitter or to the MS location directly. Path loss models of radio signals can be used to translate RSS measurements to distance estimates between the receiver and APs, from which the MT position can be obtained using trilateration process [1],[3]. However, for indoor positioning, non-line-of-sight propagation and attenuations caused by walls, other structures, and even people cause significant fluctuations to RSS, which makes the simple path loss models too inaccurate in many real life situations [4].

Fingerprinting approaches are based on experimental models that relate the measured RSS values directly to the measurement position; these models are called radio maps. The estimation of a radio map is based on data collected off-line from several locations (calibration points) covering the area where positioning service is performed.

Compared to other RSS-based methods, fingerprinting algorithms are more robust against the signal propagation fluctuations generated by environment characteristics, since fingerprinting makes use of the location-dependent variation of RSS. As an example, averages of RSS from one AP, measured in several locations, are shown in Fig. 1.

In indoor positioning with WLAN fingerprints, the positioning accuracy is affected by the amount of RSS data details that are stored into the radio map. Previously has been found that the calibration point (CP) grid density has an effect on the accuracy [5]; the mean positioning error grows gradually but clearly as the distance between calibration points grows. It has also been found that when probabilistic methods are used for position estimation, and the RSS distribution in calibration point is approximated using histograms, the number of histogram bins affects positioning accuracy, as well as the way how the information about the measurement direction is treated in the radio map [6].

However, due to the radio map configurations chosen for

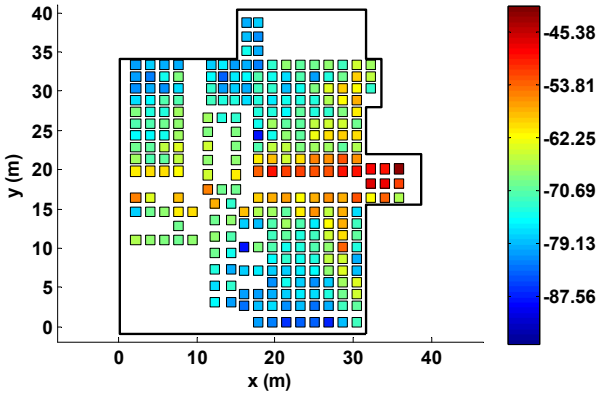

Figure 1. Average RSS of one AP measured on calibration points.

The research work presented in this paper has been funded in part by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°227890 (GRAMMAR project) and the Finnish Funding Agency for Technology and Innovation under funding decision 40200/07 (3D-NAVI_EXPO).

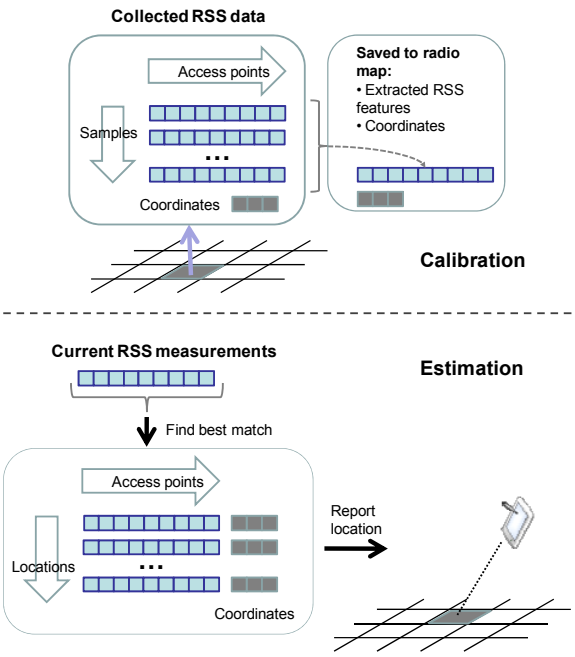

Figure 2. Calibration and estimation phases of fingerprinting based positioning.

comparisons in [6], answers to some questions still remained unclear as listed in the following:

- What is the number of histogram bins after which the positioning accuracy does not improve any more even if the number of bins is still increased?
- Is it beneficial to have a separate bin for missing samples, i.e. samples with measurement value below the minimum of correct RSS measurement range?
- Does it improve the positioning accuracy if the information of the measurement directions is included into the radio map?

We addressed these questions by using the same WLAN RSS data as was used in [6]. From these data, we computed several radio maps from the CP measurements, used independent test point data to estimate positions with the different radio maps, and compared the positioning accuracies obtained using these radio maps.

II. FINGERPRINTING

WLAN fingerprinting consists of two phases: in calibration phase, off-line-collected RSS data is used to generate a radio map, and in estimation phase new RSS measurement vectors are related with the information stored in radio map. This process is illustrated in Fig. 2. For each CP, the radio map contains the known coordinates (or other suitable location identifier) of the CP together with RSS features extracted from RSS measurements collected in the CP. The CP coordinates together with the extracted features is called a fingerprint.

The RSS features reported in the literature to have been used for positioning with WLAN fingerprints include the sample mean of RSS measurements [4],[7],[6] and approximations of the probability density functions (PDF) of the RSS samples [8],[9],[5],[6],[10]. In each CP, the sample means or PDFs are computed separately for each AP.

III. ALGORITHMS

In this section, we present the algorithms used in this paper for fingerprinting based positioning. The simplest of the algorithms is pattern recognition based on RSS sample means. Maximum Likelihood (ML) estimation and Minimization of Expected Error (MEE) are probabilistic algorithms, where histograms of measured RSS can be used to approximate their PDFs

A. Pattern Recognition

For pattern recognition, the calibration measurements from the CPs are collected into vectors. The vector elements are the RSS measurements associated with different MAC addresses, so that the order of MAC addresses is fixed. In calibration phase, the RSS values $\zeta_{i,j}$ measured in location ℓ are collected to a matrix

$$Z_\ell = \begin{bmatrix} \zeta_{1,1} & \cdots & \zeta_{1,n_{AP}} \\ \vdots & \ddots & \vdots \\ \zeta_{n_S,1} & \cdots & \zeta_{n_S,n_{AP}} \end{bmatrix}, \quad (1)$$

where n_S is the number of measurement vector samples in location ℓ and n_{AP} is the number of APs. The pattern vectors $z_\ell = [z_{\ell,1},\ldots,z_{\ell,n_{AP}}]$ for locations $\ell = 1,\ldots,n_{CP}$ are obtained by taking an average over the calibration data columns:

$$z_{\ell,j} = \frac{1}{n_S}\sum_{i=1}^{n_S}\zeta_{i,j}\ . \quad (2)$$

In estimation phase, new measurement vectors are compared with pattern vectors stored in the radio map. The comparison is based on distances in signal space [7]. A simple choice for distance measure between a measurement vector $y = [y_1,\ldots,y_{n_{AP}}]$ and the pattern vector of radio map entry z_ℓ is the usual Euclidian distance:

$$d(y,z_\ell) = \sqrt{\sum_{j=1}^{n_{AP}}(y_j - z_{\ell,j})^2}\ . \quad (3)$$

Now the position estimate can be found by searching the nearest neighbor in signal space, i.e., the pattern vector, which minimizes the distance:

$$\hat{l} = \arg\min_\ell (d(y,z_\ell))\,. \quad (4)$$

B. Maximum Likelihood

In histogram based approximation of PDF, the signal range is divided into n_B bins when the continuous or fine-resolution discrete scale becomes discrete scale with coarse resolution. The value n_B is a design parameter of an algorithm; it has an effect to the obtainable positioning accuracy, memory requirement of the radio map, and computational load of position estimation. A radio map with n_B bins requires memory for $n_{CP} n_{AP} n_B$ elements.

A histogram based radio map stores marginal distributions $p(y \mid \ell)$ for each CP location ℓ, i.e. the conditional probabilities that the measured RSS vector y can be observed at location ℓ. In estimation phase, the conditional probabilities are employed to calculate the posterior probability $p(\ell \mid y)$, i.e. the probability of being located at ℓ, given the measured RSS values y.

The posterior probabilities of the locations can be computed using Bayes' Theorem [8]:

$$p(\ell \mid y) = \frac{p(y \mid \ell)\, p(\ell)}{p(y)} = \frac{p(y \mid \ell)\, p(\ell)}{\sum_{\ell' \in \mathscr{L}} p(y \mid \ell')\, p(\ell')} \quad (5)$$

where $p(\ell)$ is the prior probability of being at location ℓ. For snap-shot type of positioning, where prior information of location is not available or not used, a non-informative uniform distribution is used. The set $\mathscr{L}$ contains all the possible locations and $p(y)$ is the probability of the measurement vector $y = [y_1, \ldots, y_{n_{AP}}]$ over all locations; $p(y)$ does not depend on location and can be treated as normalizing constant.

The maximum likelihood estimate of the posterior probability is given as:

$$\hat{\ell} = \arg\max_{\ell} p(\ell \mid y). \quad (6)$$

Using Bayes' theorem (5), assuming non-informative prior (equal $p(\ell)$ for all ℓ), noting that $p(y)$ does not depend on ℓ, and assuming independence of observations y_i, i.e., $P(Y_1 = y_1, \ldots, Y_{n_{AP}} = y_{n_{AP}}) = \prod_{i=1}^{n_{AP}} P(Y_i = y_i)$, the ML estimate (6) can be computed using the following [9]:

$$\hat{\ell} = \arg\max_{\ell} \left(\prod_{i=1}^{n_{AP}} p(y_i \mid \ell) \right). \quad (7)$$

C. Minimization of Expected Error

The MEE estimate

$$\hat{\ell} = E(\ell \mid y) = \sum_{\ell' \in \mathscr{L}} \ell'\, p(\ell' \mid y) \quad (8)$$

minimizes the expected squared location error, if the location variable is numerical [8], which is the case when location variable is coordinates. Assuming equal $p(\ell)$ for all ℓ and independence of observations y_i, the posterior probabilities $p(\ell \mid y)$ in (8) can be computed using Bayes' theorem:

$$p(\ell \mid y) = \frac{\left(\prod_{i=1}^{n_{AP}} p(y_i \mid \ell) \right)}{\sum_{\ell' \in \mathscr{L}} \left(\prod_{i=1}^{n_{AP}} p(y_i \mid \ell') \right)}. \quad (9)$$

Compared to the ML algorithm, this algorithm has the advantage that it allows the estimate to interpolate between the CPs. Its disadvantage is that the nominator in (9) needs to be evaluated, which increases the computational load in the estimation phase.

IV. Data and Experiment Setup

The experiments were carried out using the same WLAN RSS data as in [6]. The data was collected at Tampere University of Technology (TUT) library with a mobile terminal (Nokia N800 Internet Tablet) and it contains measurements from 287 CPs and 77 test points (TP). Calibration data were measured from four directions in each CP; on average it contained 13 samples from each direction. In test points, the data were collected from only one direction that varied randomly from point to point, and the average number of samples from each test point was 10. In total, the calibration and test data sets contain 15372 and 791 RSS vector samples. The test area and the CP and TP locations are shown in Fig. 3. The separation between the CPs varied between 1.5 and 2.1 m.

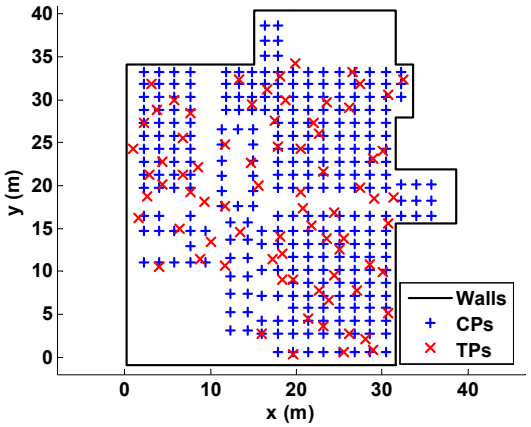

Figure 3. Floor plan and locations of calibration and test points.

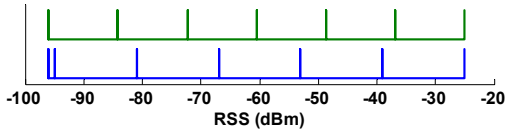

Figure 4. Edges of 6 histogram bins, evenly and unevenly distributed.

To compare the effects of the aforementioned design choices, i.e., number of histogram bins, separate bin for missing values, direction information in radio map, and combining of the signals from adjacent transmitters, we varied them to compute several different radio maps from the CP data. Then the test point data was used to estimate positions with these radio maps, and finally the positioning accuracies obtained with the different radio maps were compared with each other.

A. *Effect of Bin configuration*

To study the effect of the number of bins, radio maps were generated with 2-15 bins. To investigate the effect of separate bin for missing samples, we performed the number of bins test in two different ways. For each number of bins, there was one bin configuration with even bin widths where the RSS minimum (-96 dBm used to indicate the missing samples) was classified to the same bin with other small values. Another configuration was defined to have one narrow bin for the minimum RSS value while the rest of the bins had equal widths. An example of the used bin widths for six bins is shown in Fig. 4.

B. *Benefit of Direction Information in Radio Map*

The effect of direction information in radio map was tested by generating three radio maps where the information on measurement direction was treated differently. One radio map (*rmap1*) included the direction information such that separate fingerprints (radio map entries) were generated for each measurement direction in each CP. In the second radio map (*rmap2*), the direction information was omitted, thus all the measurements from the same CP regardless of the measurement direction were combined to one fingerprint. The resulting radio map had a size of only a fourth of size of the first radio map. On the other hand, now there were four times more calibration data available for each radio map entry.

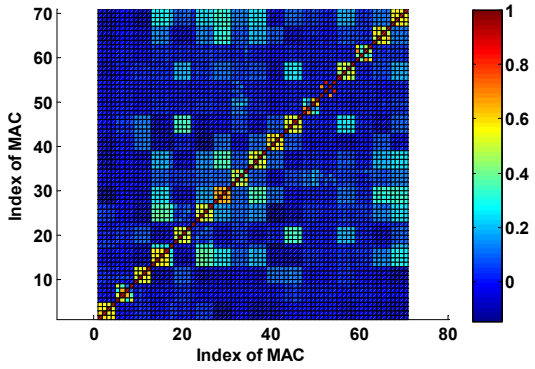

Figure 5. Correlation coefficients between the measurements from different MACs.

To analyze the effect of combining samples from all directions in isolation from the sample size for a fingerprint, we created the third radio map (*rmap3*) where only one fourth of the samples from each direction were combined together to compute one fingerprint. This yields approximately the same calibration set size for each fingerprint as was available when each direction was treated separately.

C. *Effect of Combining Measurements from Adjacent Transmitters*

In our test area, the infrastructure WLAN access points contain four antennas each. Therefore, the measured data contains groups of four MAC-addresses, where the RSS measurements within each group show high correlation due to the mutual proximity of their sources. Fig. 5 shows the correlation coefficients computed for the RSS data from different MACs where the higher correlations within the groups of four MACs can clearly be observed. On the other hand, there is quite large amount of missing samples in

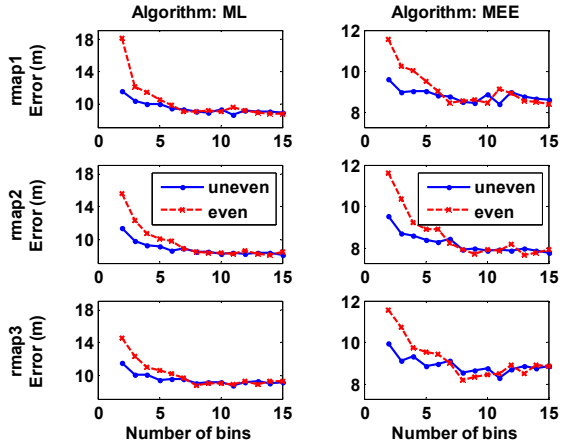

Figure 6. Root Mean Square Error with different bin configurations.

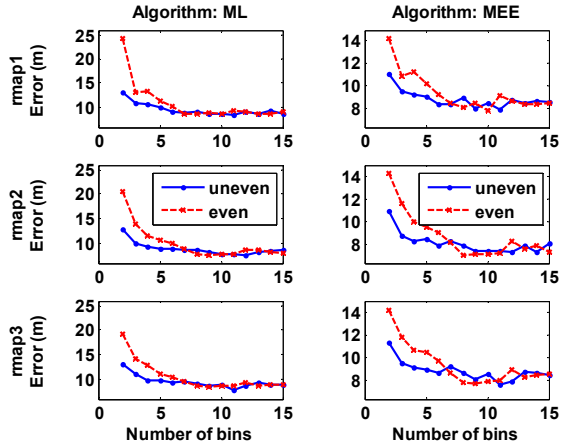

Figure 7. 75th percentile of positioning error.

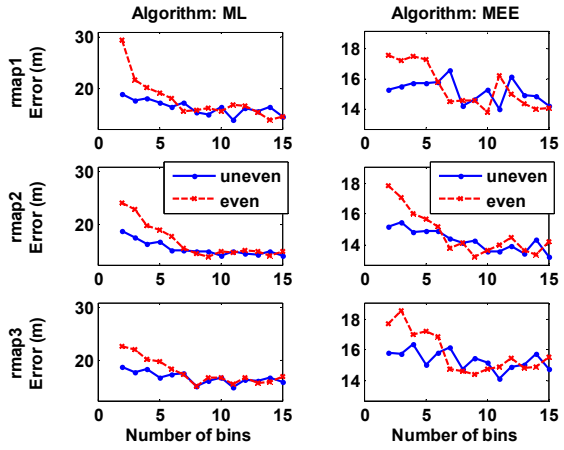

Figure 8. 90th percentile of positioning error.

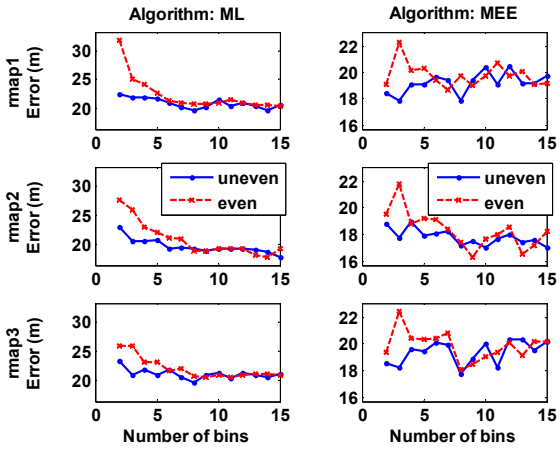

Figure 9. 95th percentile of positioning error.

measurement sets [10]. One approach for mitigating the problem of missing samples is to combine the measurements from the correlating sources before computations, both in radio map and in estimation. We tried five alternative methods for combinations: mean, median and maximum of the four signals from the same AP, and the mean and median where minimum values were excluded from the data.

V. RESULTS

For each number of bins (uneven and even bin widths), three different treatments of the direction information a radio map was generated, and they were used with the test data to estimate positions using the three positioning algorithms described earlier.

A. Effect of Bin configuration

The results of the bin configuration tests are shown in Figs. 6 -10 where RMS errors, 75th, 90th, and 95th percentiles of positioning errors and maximum positioning errors are plotted for all the test cases. The results are plotted for all the three treatments of direction information (*rmap1*, *rmap2*, *rmap3*) to illustrate the fact that the bin configuration results do not depend on how the direction information is used.

According to the test results, the RMS position error as well as the 75th and 90th percentiles of position error distance decrease as the number of bins increases until it gets the value seven or eight, after which the error measures stop decreasing (Figs. 6 -8). The number of bins has very little impact, if any, on the maximum errors (Fig. 10). From the plots of the 95th error percentile it can be observed that with ML algorithm, the error measure clearly drops as a function of the number of bins, but the effect is not so significant with MEE algorithm (Fig. 9).

The results in Figs. 6 -10 also show that it depends on the number of bins whether a separate bin for missing samples brings a benefit or not: if the number of bins is six or less, the positioning errors are smaller with uneven bin widths where there is a separate bin for RSS minimum. With numbers of bins seven or more, it seems to vary randomly whether the positioning accuracy is better using unequal or equal bin widths. The effect of the number of bins on the accuracy is more significant with even bin widths, as the errors with fewer

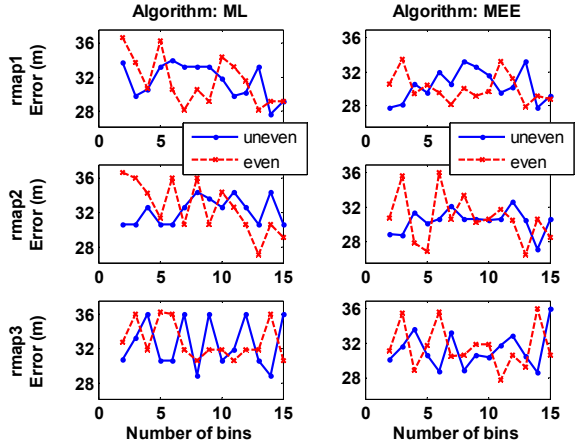

Figure 10. Maximum positioning error.

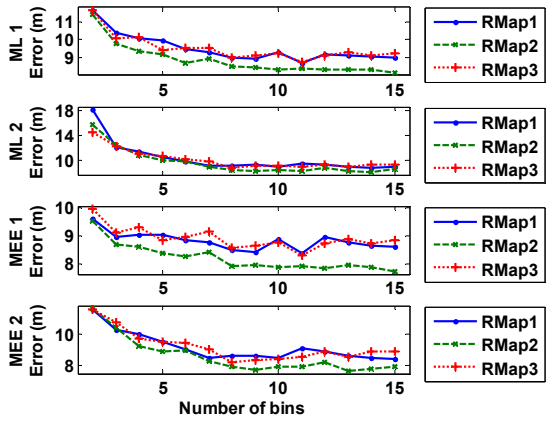

Figure 11. Direction Information in Radio Map, Root Mean Square Error.

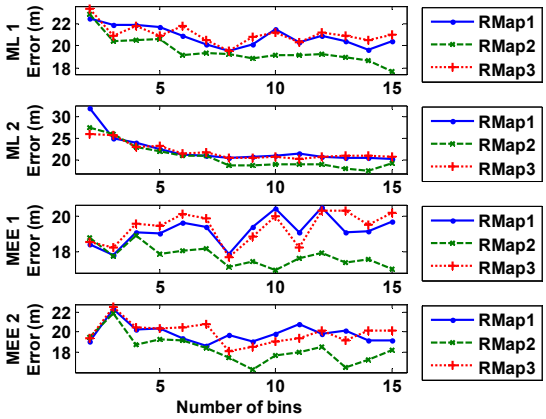

Figure 12. 95th percentile of positioning error.

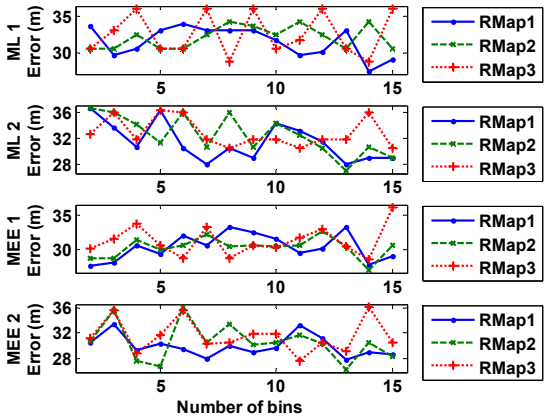

Figure 13. Maximum positioning error.

bins are larger with this configuration.

B. *Benefit of Direction Information in Radio Map*

The test results using radio maps where the direction information is treated differently are shown in Figs. 11 -13. In the figures, results are shown for the both histogram based algorithms (ML and MEE), and for configurations with uneven (ML1, MEE1) and even bin widths (ML2, MEE2).

In the tests three radio maps were compared:

- *rmap1*: Separate fingerprint for each direction, therefore less data per one fingerprint than in *rmap2*,
- *rmap2*: CP data from all directions lumped to one fingerprint, therefore containing the largest amount of data per fingerprint, and
- *rmap3*: CP data from all directions lumped to one fingerprint, but only 25% of the data used for radio map, therefore contains the same amount of data per fingerprint as *rmap1*.

The RMS error and 95th error percentile are presented in Figs. 11 -12. As expected, based on the sample size, the performance of *rmap2* is always better than the performance of *rmap1*. The performance of *rmap3* is comparable with *rmap2*; which one is better, varies with the number of bins, but variation appears to be random. Hence the crucial factor to explain the differences between the radio maps seems to be the sample size used to compute one fingerprint.

As with bin configuration tests, neither treatment of the direction information seems to have effect on maximum position error (Fig. 13); here the superiority of radio maps seems also to behave randomly as the number of bins changes. The effect is similar regardless of the positioning algorithm and the bin configurations.

The effect of different treatments of direction information in pattern matching with sample means is illustrated in Fig. 14. The difference between the cumulative distribution functions of the position errors is fairly small: the difference between the medians obtained using the different radio maps is less than 0.5 m. With small and large error percentiles the curves mostly overlap. The RMS errors using the radio maps are 10.84, 10.79, and 10.95, which do not either show significant differences.

An interesting observation can be made by comparing the RMS errors of pattern matching with RMSE results of MEE algorithm shown in Fig. 6. With only two unevenly distributed bins, the RMSE values of MEE algorithm are bellow 10 m, slightly smaller than errors obtained using pattern matching with sample means. This two-bin configuration is actually an experimental coverage map of the area, containing only probabilities for APs to be hearable in CPs. In this case, probabilities of RSS belonging to these two bins can be expressed using one parameter only, as the other probability can be obtained by subtracting the first from the probability 1. Therefore, we have two radio maps with only one feature representing each CP-MAC pairs, where the experimental coverage map with probabilistic MEE algorithm provides

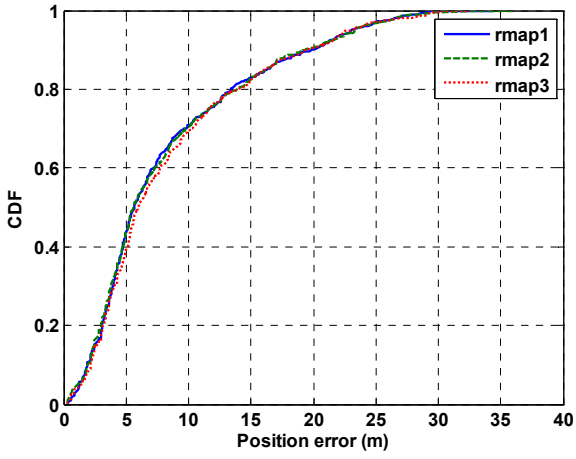

Figure 14. Effect of direction information in radio maps for pattern matching with sample means.

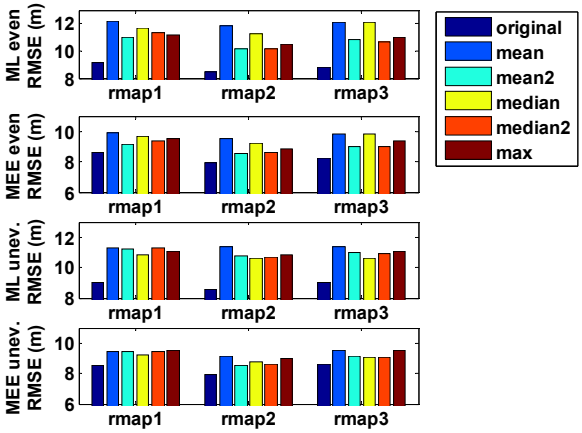

Figure 15. Effect of combining signals from correlating sources for ML and MEE algorithms with 8 evenly and unevenly distributed bins.

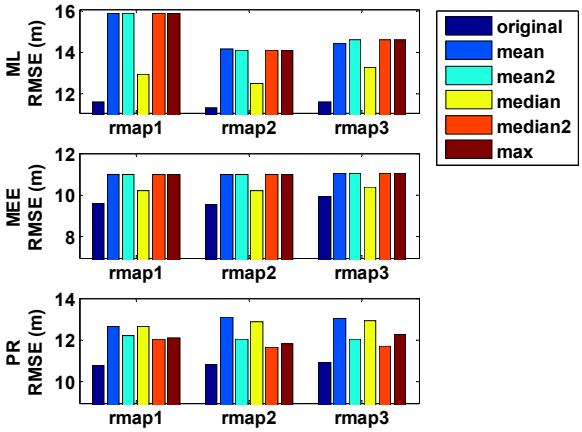

Figure 16. Effect of combining signals from correlating sources for ML and MEE algorithms with 2 unevenly distributed bins and sample mean based pattern matching.

better accuracy than pattern recognition with sample mean.

C. *Effect of Combining Measurements from Adjacent Transmitters*

To mitigate the problem of missing samples, and to decrease the size of the radio map, we tried to utilize the correlations among RSS measurements from adjacent emitters. Five alternative methods were used to combine the four signals from the same AP:

- *mean*: Mean of the signals,
- *mean2*: Mean of the signals where minimum values were excluded from the data,
- *median*: Median of the signals,
- *median2*: Median of the signals where minimum values were excluded from the data, and
- *max*: Maximum of the signals.

Positioning estimates using these combinations were computed for three bin configurations and sample mean based pattern recognition. The results of combining signals together with the original results without source combinations are shown in Figs. 15 and 16.

The RMS positioning errors for eight bins with even and uneven bin widths are shown in Fig. 15. In Fig. 16, the results for two uneven bins (experimentally determined coverage map) and sample mean based pattern recognition are shown. From the figures it is clear that the original approach without combinations is the most accurate, yielding always smaller position error than the combined versions. This can be observed with all three radio maps, all three algorithms, and all three bin configurations.

With the bin configurations where missing samples are classified to a bin separate from other measurements, shown in the two lower plots in Fig. 15 and two upper plots in Fig. 16, the combination method that yields the best accuracy is the median of all the four correlating signals. Using probabilistic algorithms with eight evenly distributed bins and with pattern recognition based on sample mean, the accuracy using combined signals is the best with either mean or median while excluding missing samples, i.e., taking into account only the samples larger than the minimum value. (Fig. 15, two upper plots, and Fig. 16, the lowest plot). This is reasonable, as these algorithms do not include inherent mechanism for handling the missing samples, which is the case when using the histogram based algorithms with separate bins for missing samples.

VI. Conclusions

In this paper, questions related to design of WLAN radio map for fingerprinting based positioning were investigated.

The results of our experiments show that with histogram based algorithms, the positioning error decreases as the number of histogram bins increases, until the number of bins reaches seven or eight, after which the error measures stop decreasing. With the numbers of bins lower than these, the uneven bin distribution with separate bin for missing samples gives better accuracy than even bin widths with missing samples included to the bin for measurements with lowest signal strength. With only two unevenly distributed bins, one for missing samples and the other for all the hearable signals, the radio map is actually an experimental coverage map. The number of parameters required for this is the same as for pattern matching based on sample mean, but better accuracy is obtained using MEE algorithm and experimental coverage map.

In tests regarding the effect of direction information in radio map, if the sample size for one fingerprint is same in both cases, the positioning performances are about the same with a radio map where separate fingerprints are generated for each measurement direction compared to a radio map where calibration samples from all the four measurement directions are included into one fingerprint. Hence the crucial factor to explain the differences between the radio maps is the sample size used to compute one fingerprint. Therefore, the same

accuracy can be obtained using less calibration samples, if samples from all directions are lumped into one fingerprint; this decreases also the size of the radio map and computational load of the positioning algorithm.

From the test results regarding the combination of signals from adjacent, correlating sources, it is obvious that combining the signals does not improve positioning accuracy when compared with the case where each signal was treated separately. However, the size of the radio map with combined signals is smaller than without combination, while the accuracy degradation with the best combination methods is only 1-2 m, depending on the positioning algorithm and histogram bin configuration. For radio maps where missing samples are treated separately, the best method for signal combination was the median. If the missing samples are not treated separately in the radio map, the best methods for combining the signals were either mean or median of the signals exceeding the minimum value.

REFERENCES

[1] M. Wallbaum and S. Diepolder, "Benchmarking wireless LAN location systems," in *Proc. IEEE International Workshop on Mobile Commerce and Service*, Munich, Germany, 2005. pp. 42-51.

[2] A. Smailagic, and D. Kogan, "Location sensing and privacy in a context-aware computing environment," *IEEE Wireless Communications,* vol. 9, no. 5, pp. 10-17, October 2002.

[3] A. Kotanen, M. Hännikäinen, H. Leppäkoski, and T. D. Hämäläinen, "Positioning with IEEE 802.11b Wireless LAN," in *Proc. Int. Symp. Personal, Indoor and Mobile Radio Communications,* Beijing, China, 2003.

[4] P. Bahl and V. N. Padmanabhan," Radar: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, Tel Aviv , Israel, 2000. pp. 775-785.

[5] A. Perttula, H. Leppäkoski, S. Tikkinen, and J. Takala, "WLAN positioning on mobile phone," in *Proc. IAIN World Congress*, Stockholm, Sweden, 2009.

[6] H. Leppäkoski, S. Tikkinen, A. Perttula, and J. Takala, "Comparison of Indoor Positioning Algorithms Using WLAN Fingerprints," in *Proc. ENC-GNSS 2009*, Naples, Italy, 2009.

[7] P. Prasithsangaree, P. Krishnamurthy, and P. Chrysanthis, "On indoor position location with wireless LANs," in *Proc. Int. Symp. Personal Indoor and Mobile Radio Communications*, Lisboa, Portugal, 2002. pp. 720-724.

[8] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks,* vol. 9, no. 3, pp. 155-164, July 2002.

[9] M. Youssef, A. Agrawala, and A. Shankar, "WLAN location determination via clustering and probability distributions," in *Proc. IEEE Int. Conf. Pervasive Computing and Communications*, Fort Worth, Texas, USA, 2003, pp. 143-150.

[10] H. Leppäkoski, S. Tikkinen, A. Perttula, and J. Takala, "Normalization of Signal Strength Measurements for WLAN Based Indoor Positioning," in *Proc. ION GNSS 2009*, Savannah, Georgia, USA, 2009.